Computer Networks 172 (2020) 107169

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

Price-aware traffic splitting in D2D HetNets with cost-energy-QoE tradeoffs☆

Nadine Abbas[a,*], Sanaa Sharafeddine[a], Hazem Hajj[b], Zaher Dawy[b]

[a] Department of Computer Science and Mathematics, Lebanese American University, Beirut, Lebanon
[b] Electrical and Computer Engineering Department, American University of Beirut, Beirut, Lebanon

ARTICLE INFO

Keywords:
Heterogeneous networks
Traffic splitting
Energy consumption
Monetary cost
QoE
D2D communication
Multi-RAT

ABSTRACT

With the advances in wireless technologies, there has been tremendous increase in user demand for resource-intensive mobile Internet services. This has been coupled with the limited capabilities of mobile devices and the high level of user expectations in terms of both quality of experience (QoE) and cost. An attractive enhancement technique is to utilize the co-existence of multiple interfaces in wireless devices to connect simultaneously to different access networks including cooperation over device-to-device (D2D) links. To this end, we present in this work optimized user-centric traffic splitting strategies in heterogeneous networks to achieve a high QoE level for video on demand applications with low cost and energy consumption for end users. We formulate a multi-objective optimization problem considering different pricing models with prediction whereby a user can estimate the links' performance for future time slots and make suitable decisions accordingly. We evaluate the proposed strategies and demonstrate their effectiveness using parameters determined via experimental measurements to provide an evaluation under realistic operational conditions. Results provide useful insights on the tradeoffs between energy consumption, cost and quality of experience.

1. Introduction

The vision towards future wireless networks is to optimize resource allocation by allowing the dynamic utilization of spectrum and multiple access technologies in heterogeneous networks (HetNets) to meet the tremendous traffic demands [1]. As shown in Fig. 1, HetNets include macro cells served by base stations (BS) covering large coverage areas, and small cells served by low-power access nodes or mobile terminals for device-to-device (D2D) cooperation to increase capacity in hotspots with high user demand and to fill in areas with weak coverage. These lead to heterogeneous network deployments which require mobile devices to function under seamless operation over multiple wireless interfaces simultaneously. HetNets bring several technical challenges due to the presence of multi-radio access technologies (RATs), and the decentralized deployment in licensed and unlicensed spectrum. Resources should be allocated to maximize system performance, while minimizing transmission cost and energy consumption. In addition, decisions need to adapt to the fast variation of the environment in terms of users' interface availability and channel conditions [2,3].

In conventional networks, a mobile device can use one wireless interface at a time to download data, which is denoted as network selection [4–10]. Network selection allows moderate enhancements in terms of throughput while keeping energy consumption and cost low. In general, mobile devices tend to switch to WiFi aiming at reducing the cost of using cellular networks. Some new smartphones introduced auto-switching between WiFi and cellular data networks to avoid poor and unstable WiFi connections. The utilization of multiple wireless interfaces simultaneously, which is denoted as traffic splitting, achieves higher throughput and quality of experience with a trade off in terms of cost and energy consumption [11–17]. The traffic splitting heuristic approach proposed in our previous work [18] demonstrated potential gains that could be further enhanced by optimal solutions. This motivates the development of enhancement techniques that can balance between cost, energy consumption and user experience.

In this work, we address user-centric real-time optimal traffic splitting in D2D heterogeneous networks to provide the user with a tradeoff between cost, energy consumption and user experience while using video on demand streaming applications. The main contributions are summarized as follows:

1. We address the general traffic splitting problem aiming at minimizing the cost and energy consumption while keeping the user

☆ In this work, we demonstrate that optimal traffic splitting considering D2D cooperation in heterogeneous networks provide useful insights on the tradeoffs between energy consumption, monetary cost and quality of experience, with very low solving time, while using video on demand streaming.

* Corresponding author.
E-mail addresses: nadine.abbas@lau.edu.lb (N. Abbas), sanaa.sharafeddine@lau.edu.lb (S. Sharafeddine), hh63@aub.edu.lb (H. Hajj), zd03@aub.edu.lb (Z. Dawy).

https://doi.org/10.1016/j.comnet.2020.107169
Received 2 October 2019; Received in revised form 8 February 2020; Accepted 23 February 2020
Available online 24 February 2020
1389-1286/

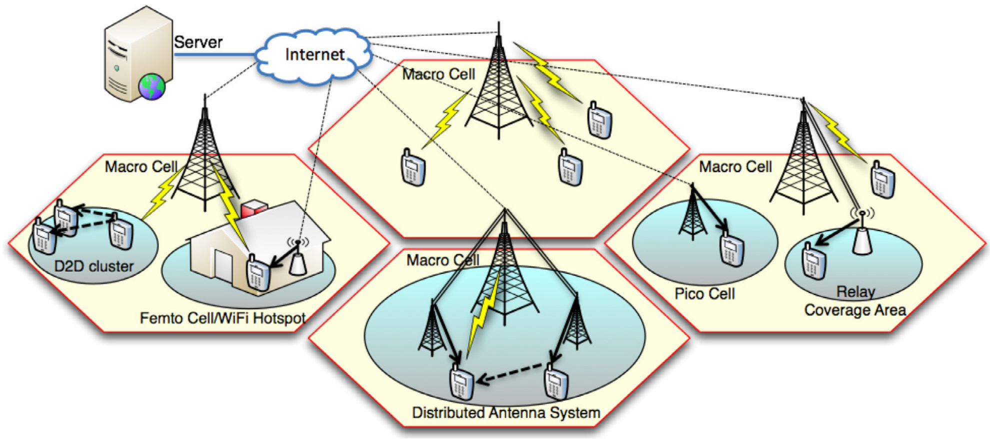


Fig. 1. Heterogeneous networks formed by macro cells, small cells and WiFi hotspots, allowing the users to communicate through one or multiple wireless interfaces including D2D cooperation.

satisfaction high. We focus on user-centric traffic splitting decisions in D2D cooperative HetNets where a user can download data not only from existing wireless networks (such as 4G/5G cellular or WiFi), but also from peer devices over D2D links. We allow the proposed approach to function in the background at the user side without any intervention from the network or the server, and without performing any changes to the technologies standards. We formulate a multi-objective optimization problem to capture cost-energy-QoE tradeoffs as a function of the dynamic variation of various system and channel parameters. We aimed at maintaining a target user satisfaction and quality level by satisfying a minimum buffering time requirement and minimizing the stalling events while using on demand video streaming applications. The problem is a mixed-integer non-linear program. To solve the problem and reduce its complexity, we transformed the problem into a mixed-integer linear program.

2. We consider different cost models including usage-based and tiered pricing. We assume cooperative pricing, which considers smart data pricing mechanisms that do not only rely on simple byte-counting schemes, but also include remaining data budget and subscription plan dynamic pricing.
3. We decide on the best download strategy considering an estimation of the dynamic system parameters variation for the next few time slots to achieve higher performance gains. Our proposed model does not only provide the user with the best performance at discrete time sample points, represented by time slots, it also estimates the network performance for multiple time slots ahead and makes decisions accordingly. We consider transmission rate estimation and overall performance of every interface for the next N time slots in our optimization problem formulation to decide on behalf of the user on the amount of data packets to be downloaded. Considering transmission rate estimation for the next few time slots will allow the user to defer the transmission at a given time slots, to another futuristic time slots where the performance of the network is better, hereby, reducing the cost and energy consumption while maintaining a target quality of experience.

This paper is organized as follows. Related work is presented in Section 2. The system model is presented in Section 3. The optimal traffic splitting approach is detailed in Section 4. Performance results are presented and explained in Section 5. Finally, conclusions are drawn in Section 6.

2. Related work

In this section, we first survey recent literature addressing network selection and traffic splitting in heterogeneous networks and then present existing related work that considers monetary cost in traffic offloading decisions.

2.1. Network selection and traffic splitting

There has been rich literature leveraging the benefits of multiple access networks for improved user experience and network performance. This literature can be broadly classified into network selection and traffic splitting, where the former selects a single network to be used by one user at a time and the former allows simultaneous use of multiple networks. Considering network selection, the authors in [4] modeled the problem as a repeated stochastic game and solved it using Lyapunov optimization algorithm to select one access network for each user at one time while maintaining a balanced load among the different access networks. The work in [5] proposed joint cell activation and selection mechanism to mitigate high energy consumption and network interference caused due to the dense existence of small cells in ultra-dense heterogeneous networks. Another related work on network selection considered a centralized resource manager to efficiently and fairly utilize the limited radio resources of available wireless networks [6]. In this context, the authors handled the problem of radio access network selection by modeling it using Semi Markovian Decision Process for sequential decision making. The authors in[7] used machine learning and game theory for online and offline user-centric network selection aiming at reducing the number of frequent switching, increasing the possibility of gainful switching, boosting user experience, and improving resource utilization. In [8], the authors proposed a hybrid unicast-multicast utility-based network selection algorithm aiming at minimizing outage percentage and energy consumption while increasing average quality of video data delivery in dense heterogeneous networks. In [9], the authors proposed an exponential utility function based on aggregate additive and multiplicative multi-criteria utility functions for network selection in heterogeneous networks. The authors is [10,11] aimed at enhancing video streaming quality of experience by considering network selection and traffic splitting while adapting the video bit rate, respectively, in HetNets. Regarding traffic splitting that concurrently delivers traffic over multiple network interfaces, the work of [12] utilized dual connectivity that allows one user equipment to simultaneously connect to a macro cell and a small cell. The work minimized the overall experienced delay by optimally splitting traffic over both connections. In [13], the authors proposed device-centric mechanisms that simultaneously utilize multiple wireless interfaces to provide optimized performance for data downloading as well as video streaming. In addition, experimental results are conducted to demonstrate the effective and practical aspects of the proposed methods. In [14], the authors devise dynamic traffic splitting techniques for video traffic over both WiFi and cellular links in a way to optimize user experience while limiting energy consumption and delay. In [15], the authors presented a mode selection technique including traffic splitting over WiFi and cellular networks using Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) considering transmission rate capacity, delay and monetary cost tradeoff. The authors in [16] proposed traffic splitting approach for proportional load balancing across two radio access networks (LTE and Wi-Fi) aiming at reducing delay while increasing average system throughput especially at high density of users. Traffic splitting has also been explored in the context of ultra-reliable low latency communication [17]. The authors utilized interface diversity to improve the resulting reliability in transmitting a given packet and studied the resulting tradeoff between reliability and latency taking into account interfaces with failure correlation.

2.2. Price-aware traffic offloading

There have been few works that address the economical implication of offloading from the perspective of either the network operator or the user. The work in [19] considered data offloading between cellular and WiFi networks and maximized user satisfaction represented in terms of allocating required network resources and user's sensitivity to data price while minimizing the incurred benefit loss of both network operators. In [20], the authors developed a network pricing model that can be used by network operators to leverage the benefits of licensed and unlicensed spectrum in handling the growing traffic demand between end users and content providers, which demonstrated significant gains in the payoff to all of network operators, content providers, and users. In [21], the authors jointly addressed price-based interference control and incentive-based offloading is modeled as a Stackelberg game and solved to offer utility gains to both cellular and D2D users. The work in [22] investigated the impact of using social relations to increase the revenue of wireless and social network operators. The problem of optimizing pricing strategy is modeled using two-stage Stackelberg game with users and coalition of operators as players. The authors in [23] optimized the pricing mechanism among cooperating users in a wireless-powered communication network where a user with better channel conditions sell its excess harvested energy to upload information of another user. In [24], the authors proposed a quality of service insured pricing strategy for traffic offloading to jointly enhance D2D performance, pricing and usersâ decision-making while considering reference point, probability distortion and risk-aversion of D2D communication. The authors in [25] proposed dynamic RAT selection and pricing considering multiple traffic classes for serving different applications running at the user equipment (UE) with proportionally fair ranking of the UEs and dynamic policy selection scheme for polling the clients of the network. In [26], the authors proposed a pre-scheduling traffic approach where requests are managed with different deadlines based on network congestion in a time-dependent pricing systems in order to the Internet service provider to reduce network congestion and accommodate more bursty traffic.

To the best of our knowledge, none of the existing literature jointly addressed the problem of traffic splitting among multiple wireless interfaces, monetary cost of content streaming, energy consumption, and quality requirement. The work in the literature either optimizes traffic distribution among multiple interfaces to improve quality and energy consumption or uses network selection to reduce monetary cost. In our work, we keep our problem general enough to accommodate the simultaneous use of multiple network interfaces on one device as well as device-to-device cooperation to guarantee a desired quality level while minimizing the total network cost of streaming video content and resulting energy consumption. We capitalize on predicting network conditions of the various available networks to further improve the expected gains in cost and energy savings.

3. System model

We address traffic splitting in D2D cooperative heterogeneous networks where a user can use multiple wireless interfaces to download data, either from the base stations (BSs) and access points (APs) over long range (LR) wireless technologies (such as WiFi, LTE, or 5G), or from other mobile terminals (MT) using short range (SR) wireless technologies (such as LTE-Direct, WiFi-Direct, Bluetooth or WiFi).

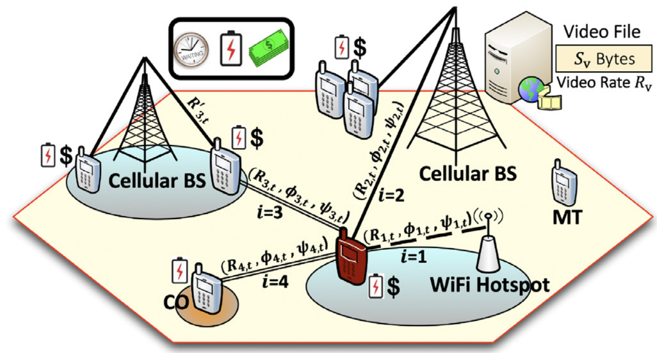


Fig. 2. D2D cooperative network formed by two cellular base stations, one WiFi hotspot, one content owner mobile terminal, and other mobile terminals with multiple wireless interfaces connections.

As shown in the example scenario in Fig. 2, the network is formed by BSs, WiFi hotspot, MTs and a content owner (CO) which is a mobile terminal that has already the data content cached. A mobile device can then use multiple wireless interfaces to download data simultaneously from BSs, APs, COs and other peer mobile terminals while using video on-demand streaming. The main aim is to reduce battery consumption and monetary cost while maintaining low delay, and desired quality of experience. We assume the availability of I wireless interfaces to download data for on-demand video streaming with a specific video bit rate R_v. We assume the video is divided into packets with a fixed size of S_p bits. Every interface i is characterized by a transmission rate $R_{i,t}$, monetary cost $\phi_{i,t}$, and energy consumption $\psi_{i,t}$ at time slot t. As shown in Fig. 2, a mobile terminal can download data using four ($I = 4$) different interfaces simultaneously, (1) WiFi ($i = 1$), (2) cellular ($i = 2$), (3) Bluetooth ($i = 3$) from a peer master mobile terminal downloading its data from a cellular base station, and (4) Bluetooth ($i = 4$) from a content owner having the data cached. When using device-to-device communication, a mobile terminal can use SR connectivity to download the desired data from (1) a content owner, or (2) a peer master device. In the first case, MT0 receives the data from the content owner ($i = 4$), which has the data cached. Accordingly, the MT of interest can directly download the data from the CO over SR connectivity such as Bluetooth without using long range connectivity such as cellular technologies. The content owner will not charge the MT for data transmission, however, the transmit energy spent by the CO and the received energy spent by the MT of interest for using SR connectivity are considered. In the second case, MT downloads data from a peer master device ($i = 3$), which does not have the data cached. Accordingly, the master device needs to download the desired data using LR connectivity from a cellular base station, and then forwards it to the MT of interest via SR connectivity such as Bluetooth ($i = 3$). The monetary cost, the transmit and receive energy consumed by the peer master mobile terminal are considered, in addition to the energy consumed by the MT of interest for downloading the data via SR connectivity. An application at the user end will decide on the number of data packets $D_{i,t}$ to be downloaded over each interface i every time slot t with duration T_s. In our work, the notion of time slots is introduced to handle discretization of the real-time aspect of the system and the operation of the proposed approach. It provides practical feasibility to make decisions periodically every time slot based on data collected and previous actions.

The video data downloaded is stored in the playing queue Q_t every time slot t. The player fetches video frames from the buffer at a constant speed defined by the video bit rate R_v. If the download bit rate is lower than the video arrival rate, the playing buffer becomes empty, and the user experiences stalling events. In this case, the player pauses,

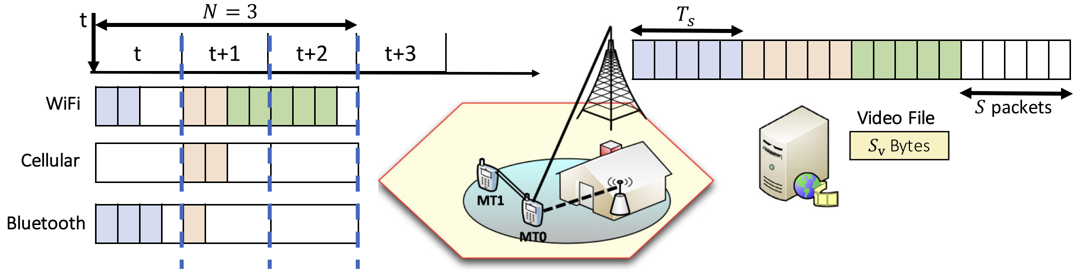


Fig. 3. Video traffic splitting over WiFi, cellular network and Bluetooth considering $N = 3$ time slots ahead.

Table 1
Main parameters and variables.

Parameters	
I	Number of available interfaces
T_s	Time slot duration
N	Number of time slots for future prediction
R_v	Video bit rate
S_p	Packet size
T_m	Minimum required of buffered video seconds
D_m	Minimum required of buffered video packets $D_m = \lceil \frac{T_m R_v}{S_p} \rceil$
i	Interface index with $i \in [1, ..., I]$
j	Time slot index with $j \in [t, ..., t+N-1]$
$R_{i,j}$	Estimated transmission rate at time slot j over interface i
$\mathfrak{C}_{i,j}$	Maximum allowed number of packets to be transmitted over interface i at time slot j
$\mathfrak{C}_j$	Maximum allowed number of packets to be transmitted over all interfaces I at time slot j
$\phi_{i,j}$	Monetary cost per data packet at time slot j over interface i
Φ_{max}	Maximum monetary cost spent over N time slots
ϕ_i	Monetary cost spent over interface i for N time slots
Φ	Total monetary cost spent over I interfaces for N time slots
$\Psi_{i,j}$	Energy consumption at time slot j over interface i
P_i	Power consumption for receiving over interface i
Ψ	Total energy consumption over all interfaces for N time slots
Ψ_{max}	Maximum energy consumption over all interfaces for N time slots
Q_t	Number of packets buffered at time slot t
$\mathfrak{N}_j$	Number of packets needed to be downloaded to prevent stalling at time slot j
B_i	Total data budget allowed over interface i
$\mathfrak{B}_{i,t}$	Remaining data budget over interface i at time slot t
Decision Variables	
$D_{i,j}$	Number of packets to be downloaded over every interface i at every time slot j. $\mathbf{D}$ is a matrix of size $I \times N$

and the last received frame freezes and is displayed until the data for the next frame is being downloaded, which leads to QoE level reduction. Accordingly, we aim at maximizing the user satisfaction by minimizing the stalling events, their length and frequency through satisfying a minimum buffering time T_m while minimizing monetary cost and energy consumption.

In our work, the decision takes into consideration the estimated performance of the interfaces for the next N time slots in order to achieve more significant performance gains in terms of energy consumption, monetary cost and quality of experience. Accordingly, at every time slot t, we consider a sliding window operation with width N, where the transmission rate $R_{i,j}$, download cost $\phi_{i,j}$, energy consumption $\psi_{i,j}$ over every interface i at every time slot $j \in [t, ..., t+N-1]$ should be estimated for the next N time slots. The amount of data to be downloaded over each interface for the next N time slots will be then updated based on the dynamic system variations to provide the best balance between monetary cost and energy consumption while meeting desired QoE level. In Fig. 3, a packet level illustration is presented where MT0 is streaming video data over three wireless interfaces: WiFi, cellular and Bluetooth. The sliding window considered is $N = 3$ time slots. The solution then will decide on the number of packets needed to be downloaded for the next three time slots over the three available interfaces simultaneously providing minimum energy consumption and monetary cost while meeting target QoE level. The main system parameters are summarized in Table 1.

4. Price-aware traffic splitting optimization

In this section, we present the user-centric traffic splitting optimization problem formulation considering system variables, objective and constraints. In our work, we aim at finding the amount of data to be transmitted over every wireless interface minimizing the monetary cost and energy consumption, while maintaining a target desired level of user quality of experience.

In our work, we decide on the number of packets $D_{i,t}$ to be downloaded over every interface i at time slot t. Since our proposed approach considers future interface performance estimations with a sliding window of width N time slots, the decision is made on the amount of data to be transmitted for the next N time slots. Accordingly, the decision variable $\mathbf{D}$ is a matrix of size $I \times N$. The traffic splitting problem is subjected to several constraints in terms of capacity limitations. The problem can be formulated as follows:

$$\underset{\mathbf{D}}{\operatorname{argmin}} \quad \alpha \frac{\Phi}{\Phi_{max}} + (1-\alpha) \frac{\Psi}{\Psi_{max}} \tag{1}$$

subject to

$$D_{i,j} \leq \mathfrak{C}_{i,j}, \forall i \in [1, ..., I], \forall j \in [t, ..., t+N-1] \tag{2}$$

$$\sum_{k=1}^{j} \sum_{i=1}^{I} D_{i,k} \geq \min\left(\mathfrak{C}_j, \mathfrak{N}_j\right), \forall j \in [t, ..., t+N-1] \tag{3}$$

- Eq. (1) is the objective function which aims at minimizing the total monetary cost Φ while keeping the energy consumption Ψ low, over the next N time slots. α is a weighting factor, varying between 0 and 1, indicating the tradeoff between cost and energy consumption. In our objective function, we used metric normalization in order to adjust the total monetary cost Φ and the energy consumption Ψ values measured on different scales (USD and Joules, respectively) to a common scale ranging between 0 and 1. $\Phi_{\max}$ and $\Psi_{\max}$ are the maximum cost and energy that may be consumed, respectively, to download the needed data over the next N time slots while using all the interfaces simultaneously. The minimum energy and monetary cost consumption is assumed to be 0 when no transmission occurs. The expressions for Φ, $\Phi_{\max}$, Ψ and $\Psi_{\max}$ are detailed in the next subsections. The value of the objective function reflects the performance of the solution in minimizing the monetary cost and energy consumption compared to the maximum use of all the interfaces simultaneously. Its value is dependent on the amount of data requested to be downloaded, which is constrained by the minimum required buffering time, the buffering queue size and the link transmission capacities in (3) to maintain the target quality of experience while using on demand video streaming. The objective function value will then range between 0 when no transmission occurs, and 1 when all the interfaces are used simultaneously.
- The first constraint (2) guarantees that the number of packets to be transmitted over interface i at every time slot j is less or equal to the interface packet capacity $\mathfrak{C}_{i,j}$, expressed as follows:

$$\mathfrak{C}_{i,j} = \frac{R_{i,j} \cdot T_s}{S_p} \tag{4}$$

- The second constraint (3) ensures that the requested data packets should arrive at the user end before a buffer underflow occurs. For video on demand streaming, synchronized media streams are played continuously while being downloaded from the application server without having to wait for the entire video to be delivered. Once the playout phase starts, the player fetches video frame from the buffer Q_t at a constant speed defined by the video bit rate R_v. Accordingly, $\lceil \frac{R_v \cdot T_s}{S_p} \rceil$ packets are needed to be played every time slot of duration T_s. When the service transmission rate is less than the video bit rate, the playing buffer becomes empty. In this case, the player pauses, and the user experiences stalling events. In order to maintain a target user satisfaction and quality level, we aimed at minimizing the stalling events by satisfying a minimum buffering time requirement T_m, which corresponds to maintaining a buffering queue size greater than D_m video data packets. Hence, the number of data packets to be downloaded at time slot j should be greater than the number of needed data packets $\mathfrak{N}_j$, estimated based on (5), for $j \in [t, \ldots, t+N-1]$, depending on the video rate R_v, time slot duration T_s, packet size S_p, data packets buffered Q_t and the minimum required buffering video data packets D_m. In case the capacity $\mathfrak{C}_j$ (6) over all the interfaces I, at time slot j, is less than the required number of data packets, the device needs to fully utilize all the interfaces capacity simultaneously to guarantee the highest quality experience possible.

$$\mathfrak{N}_j = D_m + (j - t + 1) \cdot \left\lceil \frac{R_v \cdot T_s}{S_p} \right\rceil - Q_t \tag{5}$$

$$\mathfrak{C}_j = \sum_{k=1}^{j} \sum_{i=1}^{I} \frac{R_{i,k} \cdot T_s}{S_p} \tag{6}$$

4.1. Cost minimization

We consider Φ the total transmission cost which the user pays to download data over I interfaces during N time slots. Accordingly, the total cost can be expressed as follows:

$$\Phi = \sum_{i=1}^{I} \phi_i \tag{7}$$

where ϕ_i is the cost spent when using interface i. In general, some interfaces have much higher cost than others, e.g., cellular typically has notably higher cost than WiFi or D2D over short range links such as Bluetooth. In our work, we differentiate between two cost models: (1) usage-based pricing, and (2) tiered pricing [27,28].

4.1.1. Usage-based pricing model

Usage-based pricing charges users in proportion to the amount of data consumed. In our work, we assume ϕ_i is the monetary cost per packet spent at time slot j over interface i. Accordingly, the cost ϕ_i spent by interface i using usage-based pricing model over N time slots can be represented by $\phi_{U,i}$ and expressed as follows:

$$\phi_i = \phi_{U,i} = \sum_{j=t}^{t+N-1} \phi_{i,j} \cdot D_{i,j} \tag{8}$$

4.1.2. Tiered pricing model

Tiered pricing charges users a fixed amount of money Θ_i for a monthly data capacity B_i. The fixed fee covers usage up to the capacity, after which users may pay Γ_i USD per extra MByte. The cost ϕ_i spent by interface i using tiered pricing model can be represented by $\phi_{T,i}$ and expressed as follows:

$$\phi_i = \phi_{T,i} = (1 - w_{i,t}) \cdot \theta_i \cdot \sum_{j=t}^{t+N-1} D_{i,j} + w_{i,t} \cdot \Gamma_i \cdot \left(\sum_{j=t}^{t+N-1} D_{i,j} - \mathfrak{B}_{i,t} \right) + w_{i,t} \cdot \theta_i \cdot \mathfrak{B}_{i,t} \tag{9}$$

where θ_i is the cost spent per packet when downloading over interface i when data budget is not exceeded and can be computed as follows: $\theta_i = \frac{\Theta_i}{B_i}$. A new binary decision variable $w_{i,t}$ is then introduced to indicate whether the user exceeds total budget B_i over interface i at time slot t. To accommodate for this pricing model, additional constraints are also needed:

$$\mathfrak{B}_{i,t} - \sum_{j=t}^{t+N-1} D_{i,j} \le X \cdot (1 - w_{i,t}) \tag{10}$$

$$\sum_{j=t}^{t+N-1} D_{i,j} - \mathfrak{B}_{i,t} \le X \cdot w_{i,t} \tag{11}$$

where X is a large number. Accordingly, if $w_{i,t}$ is equal to 0, constraint (11) is imposed, and constraint (10) becomes unbounded. In this case, the downloaded data is below budget and the user is charged by θ_i. If $w_{i,t}$ is equal to 1, constraint (11) becomes unbounded, and constraint (10) will hold. In this case, the data exceeds the remaining budget $\mathfrak{B}_{i,t}$, therefore, $w_{i,t}$ is set to 1 and the user is charged θ_i for the remaining capacity and Γ_i for the data exceeding the budget $\mathfrak{B}_{i,t}$. The general expression of the total monetary cost spent by the user when using all the interfaces I can be expressed as follows:

$$\Phi = \sum_{i=1}^{I} v_i \left(\mu_i \cdot \phi_{U,i} + (1 - \mu_i) \cdot \phi_{T,i} \right) \tag{12}$$

where v_i is an input binary variable indicating whether the interface i charges the user or have zero cost, and μ_i is an input binary variable indicating the type of the pricing model whether it is usage-based or tiered pricing. The maximum cost over N time slots can be computed as follows:

$$\Phi_{\max} = \sum_{i=1}^{I} v_i \sum_{j=t}^{t+N-1} \mathfrak{C}_j \cdot C_{i,j} \tag{13}$$

where

$$C_{i,j} = \mu_i \cdot \phi_{i,j} + (1 - \mu_i) \cdot \max(\theta_i, \Gamma_i) \tag{14}$$

4.2. Energy minimization

We consider Ψ as the total energy consumed by the mobile device and its peer devices to receive and transmit data. Ψ_1 represents the energy consumed by the mobile terminal, denoted as MT0, to receive data over I interfaces during N time slots. Ψ_2 represents the energy consumed by peer mobile terminals to download data and forward it to MT0. Accordingly, Ψ can be expressed as follows:

$$\Psi = \Psi_1 + \Psi_2 \tag{15}$$

where

$$\Psi_1 = \sum_{i=1}^{I} \sum_{j=t}^{t+N-1} P_i \cdot \frac{D_{i,j} \cdot S_{\mathrm{P}}}{R_{i,j}} \tag{16}$$

P_i represents the power consumed by MT0 to receive over interface i. In case of D2D cooperation, MT0 downloads data from a peer MT m, which can be (1) a content owner (CO), which has the data cached, or (2) a master device, which downloads data using WiFi or cellular and forwards it to MT0. In our work, we assume the master device downloads data with a rate R'_i higher than the forward transmission rate $R_{i,j}$ to MT0 over interface i. Accordingly, MT m consumes η_i to download and transmit data simultaneously to MT0, and then, after download all the content, MT m consumes ζ_i to transmit data to MT0. In case MT m is a CO, we consider the power consumed ζ_i by MT m to transmit data to MT0.

$$\begin{aligned}\Psi_2 = &\sum_{i=1}^{I}(1-\gamma_i)\cdot\beta_i\cdot\eta_i\cdot\frac{\sum_{j=t}^{t+N-1} D_{i,j}\cdot S_{\mathrm{P}}}{R'_i} \\ &+ \sum_{i=1}^{I}\beta_i(1-\gamma_i)\cdot\zeta_i\cdot\sum_{j=t}^{t+N-1} D_{i,j}\cdot S_{\mathrm{P}}\cdot\left(\frac{1}{R_{i,j}}-\frac{1}{R'_i}\right) \\ &+ \sum_{i=1}^{I}\beta_i\cdot\gamma_i\cdot\zeta_i\cdot\sum_{j=t}^{t+N-1}\frac{D_{i,j}\cdot S_{\mathrm{P}}}{R_{i,j}}\end{aligned} \tag{17}$$

where β_i is an input binary decision variable indicating whether the interface i is a short range connectivity and used for D2D cooperation. γ_i is an input binary decision variable indicating whether the interface i is a content owner ($\gamma_i = 1$) or a master device ($\gamma_i = 0$). The maximum energy consumed can be expressed as follows:

$$\Psi_{\max} = \sum_{i=1}^{I}\sum_{j=t}^{t+N-1} P_i\cdot T_{\mathrm{s}} + \sum_{i=1}^{I}\beta_i\left[(1-\gamma_i)\cdot\eta_i + \gamma_i\cdot\zeta_i\right]\cdot T_{\mathrm{s}} \tag{18}$$

4.3. Problem linearization

The problem is a mixed-integer non-linear program. The non-linearity comes from the objective function when tiered cost model is used. To reduce the complexity of the problem, we transformed it into a mixed integer linear program.

The tiered cost model $\phi_{\mathrm{T},i}$ in (9) can be developed and expressed as follows:

$$\begin{aligned}\phi_{\mathrm{T},i} = &\ \theta_i \sum_{j=t}^{t+N-1} D_{i,j} - \left(\Gamma_i - \theta_i\right)\cdot\sum_{j=t}^{t+N-1} w_{i,t}\cdot D_{i,j} \\ &+ w_{i,t}\cdot\left(\theta_i - \Gamma_i\right)\cdot\mathfrak{B}_{i,t}\end{aligned} \tag{19}$$

To eliminate the non-linearity, the product of the two variables ($w_{i,\,t}\cdot D_{i,\,j}$) can be replaced by a new variable $Y_{i,j}$, on which several constraints are imposed. In our case, the first variable $w_{i,\,t}$ is a binary variable, and the second variable $D_{i,j}$ is an integer continuous variable ranging between 0 and the maximum link capacity $\left(R_{i,j}/S_{\mathrm{P}}\right)$. Additional constraints are then needed to force $Y_{i,j}$ to take the value of the product of the two variables ($w_i\cdot D_{i,\,j}$), as follows:

$$Y_{i,j} \le w_{i,t}\cdot\frac{R_{i,j}}{S_{\mathrm{p}}} \tag{20}$$

$$Y_{i,j} \le D_{i,j} \tag{21}$$

$$Y_{i,j} \ge D_{i,j} - \frac{R_{i,j}}{S_{\mathrm{p}}}\cdot(1 - w_{i,t}) \tag{22}$$

$$Y_{i,j} \ge 0 \tag{23}$$

Accordingly, the total cost Φ expressed in (12) can be linearized to Φ_{L} as follows:

$$\Phi_{\mathrm{L}} = \sum_{i=1}^{I} v_i\left(\mu_i\cdot\phi_{\mathrm{U},i} + (1-\mu_i)\cdot\phi_{\mathrm{TL},i}\right) \tag{24}$$

where

$$\begin{aligned}\phi_{\mathrm{TL},i} = &\ \theta_i\sum_{j=t}^{t+N-1} D_{i,j} - \left(\Gamma_i - \theta_i\right)\cdot\sum_{j=t}^{t+N-1} Y_{i,j} \\ &+ w_{i,t}\cdot\left(\theta_i - \Gamma_i\right)\cdot\mathfrak{B}_{i,t}\end{aligned} \tag{25}$$

The problem can then be formulated as follows:

$$\underset{\mathbf{D}}{\operatorname{argmin}} \quad \alpha\frac{\Phi_{\mathrm{L}}}{\Phi_{\max}} + (1-\alpha)\frac{\Psi}{\Psi_{\max}} \tag{26}$$

subject to

$$D_{i,j} \le \mathfrak{C}_{i,j}, \forall i, \forall j \tag{27}$$

$$\sum_{k=1}^{j}\sum_{i=1}^{I} D_{i,k} \ge \min\left(\mathfrak{C}_j, \mathfrak{N}_j\right), \forall j \tag{28}$$

$$\left(\mathfrak{B}_{i,t} - \sum_{j=t}^{t+N-1} D_{i,j}\right)\cdot(1-\mu_i)\cdot v_i \le X\cdot(1-w_{i,t}), \forall i \tag{29}$$

$$\left(\sum_{j=t}^{t+N-1} D_{i,j} - \mathfrak{B}_{i,t}\right)\cdot(1-\mu_i)\cdot v_i \le X\cdot w_{i,t}, \forall i \tag{30}$$

$$Y_{i,j} \le w_{i,t}\cdot\frac{R_{i,j}}{S_{\mathrm{p}}}, \forall i, \forall j \tag{31}$$

$$Y_{i,j} \le D_{i,j}, \forall i, \forall j \tag{32}$$

$$Y_{i,j} \ge \left(D_{i,j} - \frac{R_{i,j}}{S_{\mathrm{p}}}\right)\cdot(1-w_{i,t})\cdot(1-\mu_i)\cdot v_i, \forall i, \forall j \tag{33}$$

$$Y_{i,j} \ge 0, \forall i, \forall j \tag{34}$$

The problem is then a mixed-integer linear programming (MILP). The decision variables are (1) matrix $\mathbf{D}$ of size $I\times N$ where $D_{i,j}$ is an integer indicating the amount of data packets to be downloaded over every interface i at time slot j with $j = [t, t+1, \ldots, t+N-1]$, (2) vector $\mathbf{w}$ of size I where $w_{i,\,t}$ is a binary variable set to 1 when the data downloaded exceeds the data budget at time slot t, and (3) matrix $\mathbf{Y}$ of size $I\times N$ used for problem linearization where $Y_{i,j} = D_{i,j}\cdot w_{i,t}$. The number of variables is then $2I\times N + I$. Note that in real cases, for practicality, accuracy, feasibility and to achieve performance gains, the number of interfaces used simultaneously by a mobile device and the number of time slots N considered ahead for rate estimation may be limited. In order to solve our optimization problem, we used the MILP solver of MATLAB based on the branch-and-bound method. The maximum execution time taken to solve the problem with as many as 10 interfaces and N= 8 is limited to 0.06 s on a 2.4 GHz Quad-Core Intel Core i5, the fact that makes the proposed solution suitable for applications with hard time constraints.

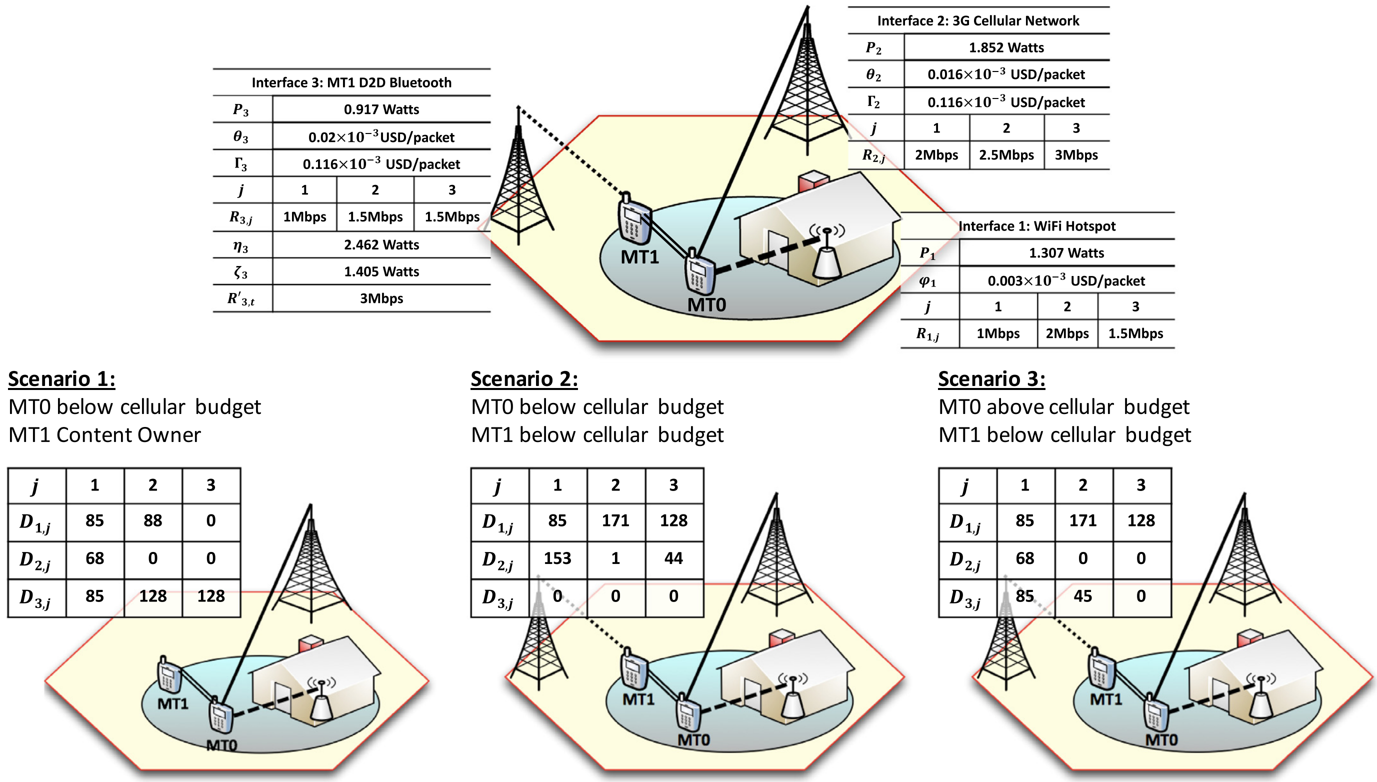


Fig. 4. Toy example illustrating optimal solutions for the proposed cost-energy-QoE tradeoff traffic splitting (TS-CEQ).

4.4. Illustration example

To illustrate an example solution for the optimization problem formulation, we consider a network composed of the user of interest MT0, cellular base stations, WiFi hotspot, and one peer device MT1 as presented in Fig. 4. MT0 aims at downloading a video with rate R_v of 2 Mbps using multiple wireless interfaces simultaneously. We assume the video data is divided into fixed size packets composed of 1460 Bytes, and the time slot duration is 1 s, which corresponds to 172 packets of video data. In our case model, we assume the data buffered Q_t is 450 packets, and the minimum required buffered data D_m is 516 packets which corresponds to 3 s of playing video data. The traffic splitting decision takes into consideration three slots ahead ($N = 3$). Accordingly, MT0 needs to download at least 238 packets in the first time slot ($j = t$) in order to maintain the minimum desired QoE level by satisfying a minimum buffering time requirement T_m as expressed in (5). In the second time slot ($j = t + 1$), MT0 needs to download at least a cumulative of 410 packets (238 packets needed in the first time slot and 172 in the second). In the third time slot ($j = t + 2$), MT0 needs to download in total 582 data packets (238 packets needed in the first time slot and 172 in the second and the third). The three interfaces characteristics are presented in Fig. 4. We assume MT0 consumes 1.307, 1.852, and 0.917 W while receiving data over WiFi, 3G cellular networks, and Bluetooth when connected to MT1, respectively. When MT0 downloads data from a content owner, the power consumed by the CO to transmit data over Bluetooth to MT0 is considered and is assumed to be 1.405 W. When MT0 downloads data from a master device, we assume the master device consumes 2.462 W to simultaneously receive data over 3G and transmit over Bluetooth, and 1.405 W to only transmit data over Bluetooth to MT0 [29–31]. WiFi network follows usage-based cost model, while 3G cellular network follows tiered-based cost model.

We consider three scenarios: (1) MT0 is below cellular budget, and MT1 is a content owner, (2) MT0 is below cellular budget, and MT1 downloads its data from a cellular base station and forwards it to MT0, and (3) MT0 is above cellular budget. In the first scenario, MT1 has the data cached, accordingly, there is no monetary cost for downloading the data from MT1 over Bluetooth. Therefore, MT0 downloads most of his data over Bluetooth to reduce the total cost, which is reflected by the optimal solution provided in Fig. 4. Using one interface WiFi or Bluetooth alone does not allow MT0 to fully download the desired data which may lead to stalling events and reduction in quality of experience. Using multiple wireless interfaces and taking advantage of D2D communication, the proposed approach was able to download the needed number of packets while maintaining the minimum quality level, while consuming 0.0016 USD and 5.46 J for one second of video streaming, which leads to an approximate consumption of 5.76 USD and 19.656 kJ for one hour of video streaming. Compared to using cellular network alone, the proposed approach was able to provide 82.74% reduction in monetary cost while consuming 6.58% more energy in one time slot with 1 s duration. In the second scenario, MT1 is no longer a CO, it downloads the requested data from a cellular base station. Since the cost of WiFi interface is lower in terms of price and power consumption, MT0 favors the use of the WiFi over cellular link. MT0 does not use D2D cooperation since the system constraints are met without the need to connect to MT1. In this case, the proposed approach consumes higher cost of 0.0043 USD and energy of 5.9 J in one second of video streaming, leading to an approximate cost of 15.48 USD and 21.24 kJ for one hour of video streaming. Compared to the use of cellular network only, the proposed approach was able to fully download the required data with 53.6% less monetary cost while consuming 15.1% more energy. In the last scenario, MT0 exceeds the cellular budget while MT1 is still below its budget. In this case, MT0 takes advantage of the WiFi network and D2D cooperation to reduce the total cost of downloading data while meeting QoE requirements. The proposed approach consumes 0.0116 USD and 8.33 J, which provides 82.75% reduction in monetary cost while consuming 62% more energy compared to using cellular network alone.

5. Performance results and analysis

To validate the proposed multi-objective traffic splitting with cost-energy-QoE tradeoffs (TS-CEQ) approach under realistic conditions, experimental measurements are used to determine WiFi and cellular key link parameters, such as effective download rate and energy consumed per second during data reception. The performance of the proposed TS-CEQ approach is quantified and analyzed for HetNet resource management.

5.1. Performance evaluation

In order to assess the performance effectiveness of the proposed TS-CEQ traffic splitting approach, we generated results for the following different strategies:

1. *WiFi only* (WO): the user downloads data using WiFi link only.
2. *Cellular only* (CeO): the user downloads data using cellular link only.
3. *Maximum rate network selection* (MaxR-NS): the link providing the higher rate is selected in every time slot.
4. *Exponential utility function for network selection* based on [9]: the user selects the link providing higher exponential utility function considering monetary cost and bandwidth tradeoff (XCB-NS). We also customized the utility function and considered two additional cases: exponential utility function for energy consumption and monetary cost tradeoff (XCE-NS), and exponential utility function considering energy consumption, monetary cost, and QoE tradeoff (XCEQ-NS).
5. *Traffic splitting over both links simultaneously* (TS-S): the user always uses both links simultaneously to download data.
6. *Traffic splitting with delay-power-QoE balance* (TS-PQ) based on [14]: the user uses one of the following strategies: (1) WiFi alone, (2) cellular alone, (3) traffic splitting, or (4) no transmission. The strategy providing higher utility function is selected on a time slot basis. The utility function is based on a Lyapunov drift-plus-penalty formulation providing delay, power and QoE tradeoff.
7. *Traffic splitting with transmission rate capacity-delay-monetary cost balance* (TS-RDC) based on [15]: the user uses one of the following strategies: (1) WiFi alone, (2) cellular alone, or (3) traffic splitting. The strategy providing higher utility function is selected on a time slot basis. The utility function is based on Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) for network selection providing transmission rate capacity, delay and monetary cost tradeoff.
8. *Traffic splitting with cost-energy-QoE balance* (TS-CEQ): the user optimally splits the data over multiple interfaces providing a balance between monetary cost, energy consumption while meeting minimum QoE requirement and taking into consideration the performance of the network within the next N time slots ahead.

5.2. Simulations setup

We evaluate the performance of our proposed approach using simulations conducted using MATLAB to stream a video using different strategies. In the simulator implementation, a mobile device downloads data for video streaming application, where the video specifications, such as video size S_{v}, duration T_{v}, and bit rate R_{v}, are obtained as input from the server before the start of the video download. The chosen video has a size of 7 MBytes, a duration of 60 s, a frame rate of 25 fps, and an arrival rate of 117 kBytes every second. Simulations are conducted for one hour of video streaming considering 60 runs of 60 s videos.

Synchronized media streams are played continuously while being downloaded from the application server without having to wait for the entire video to be delivered. The player fetches video frame from the buffer Q_t at a constant speed defined by the video bit rate R_{v}. If the download bit rate is lower than the video arrival rate, the mobile device is able to download only a fraction of the requested data. The remaining data that was not downloaded is kept to be downloaded in the next time slots. In this case, the user experiences stalling events, the player pauses, the last received frame freezes and is displayed until the data for the next frame is being downloaded. If the download bit rate is higher than the video arrival rate, the data is downloaded on time without any delay, stalls, or freeze frames.

Accordingly, at each time slot of duration T_{s} =1 s, the proposed approach makes a decision on the data to be downloaded in the next time slot taking energy consumption, monetary cost and cellular budget into consideration while meeting the minimum QoE requirement by satisfying a minimum buffering time requirement T_{m}. The decision is based on the solution of the TS-CEQ optimization problem formulated in (26)–(34) for N time slots ahead. Once the data is downloaded, the system parameters are recorded such as buffer size, downloaded data, effective transmission rate, energy consumption, QoE parameters including freeze frames, stalling length and frequency. The process is then repeated, statistics are updated every time slot until video data is completely downloaded.

In order to assess the QoE level, we used Mean Opinion Score (MOS) values expressed in Eq. (35) derived in our previous work [14] based on QoE metric presented in Recommendation ITU-T P.1201 (2013) [32]. The level of user satisfaction can be measured by MOS level ranging between 1 (bad) and 5 (excellent) [33–35]. In our work, the MOS values are estimated based on a QoE metric that is derived from standards presented in Recommendation ITU-T P.1201 (2013) considering initial buffering, and re-buffering events, their frequency and length [32]. The MOS level derived in our previous work [14] can be expressed as follows:

$$\Lambda_t = a \cdot \log(b \cdot \mathfrak{M}_t + c) \tag{35}$$

where a, b and c are found to be 0.9377, 128.9 and −427.6, respectively. $\mathfrak{M}_t$ is the QoE metric presented in [32], and can be expressed as follows:

$$\mathfrak{M}_t = 5 - \max(\min((\Omega_t + \mathfrak{T}), 4), 0) \tag{36}$$

where Ω_t and $\mathfrak{T}$ are the expected degradation caused by stalls and initial buffering till time t. They are defined as follows:

$$\Omega_t = \max(\min(s_4 + s_1 \cdot \exp((s_2 \cdot L_t + s_3) \times N_t), 4), 0) \tag{37}$$

$$\mathfrak{T} = \begin{cases} \max(\min(d_1 \cdot \log(T_0 + d_2), 4), 0), & \text{if } T_0 \geq 1 - d_2 \\ 0, & \text{otherwise} \end{cases}$$

where T_0 is the initial loading time in seconds, L_t is the averaged stalling duration in seconds and N_t is the number of stalling events excluding initial buffering. The coefficients s_1, s_2, s_3, s_4, d_1 and d_2 have the following values −1.72, −0.04, −0.36, 1.66, 0.29 and −3.29, respectively [32].

Our proposed approach is user-centric, performing autonomously, without any intervention from the network or the server, and without performing any changes to the cellular/WiFi/ Bluetooth standards. The proposed approach runs at the user side and adjusts based on what it receives from the network to provide real-time solutions for data download. The device estimates the download rate in the subsequent time slots based on the actual rates experienced in the current and the previous time slots. Accordingly, the problem does not require any resource allocation since the device makes its decisions based on the actual measured bit rates of all interfaces.

In our simulations setup, we assume the cellular and WiFi transmission rates follow the exponential distribution with different mean values as presented in the results section below. We assume the average power consumed by a mobile terminal to download data is 0.917 W over Bluetooth, 1.307 W over WiFi and 1.852 Watts over 3G cellular network. A master device consumes 2.462 W to simultaneously receive data over 3G and transmit over Bluetooth to its peer device. However, when the master device is a content owner and has the data cached, it consumes 1.405 W to transmit data over Bluetooth to its peer device [29,30].

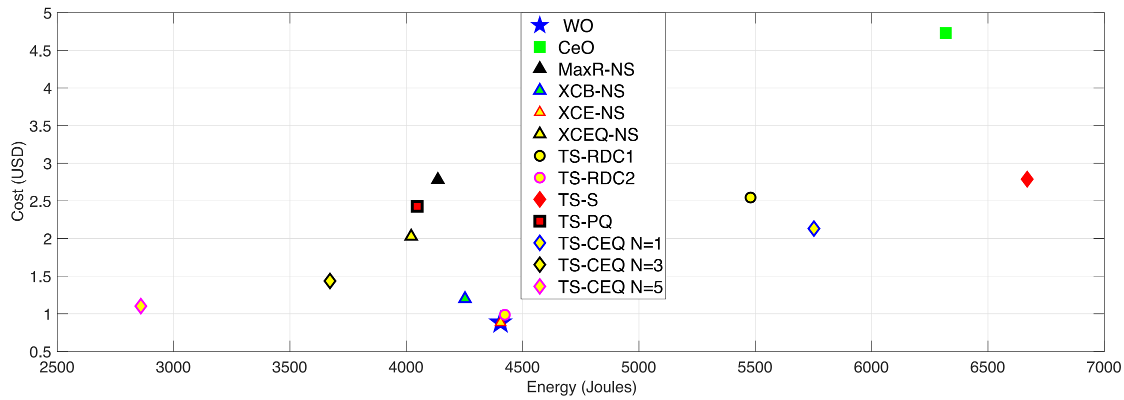


Fig. 5. Total energy expenditure (Joules) and monetary cost (USD) for every approach.

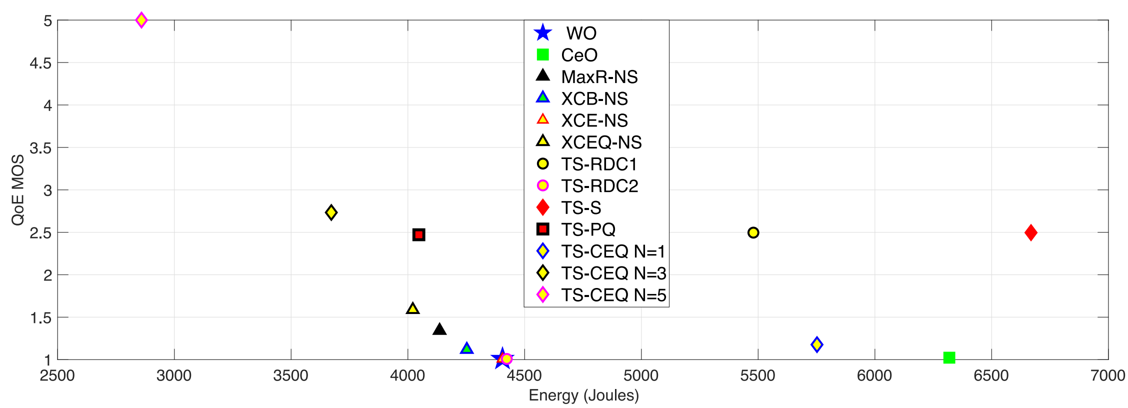


Fig. 6. Total energy expenditure (Joules) and QoE MOS for every approach.

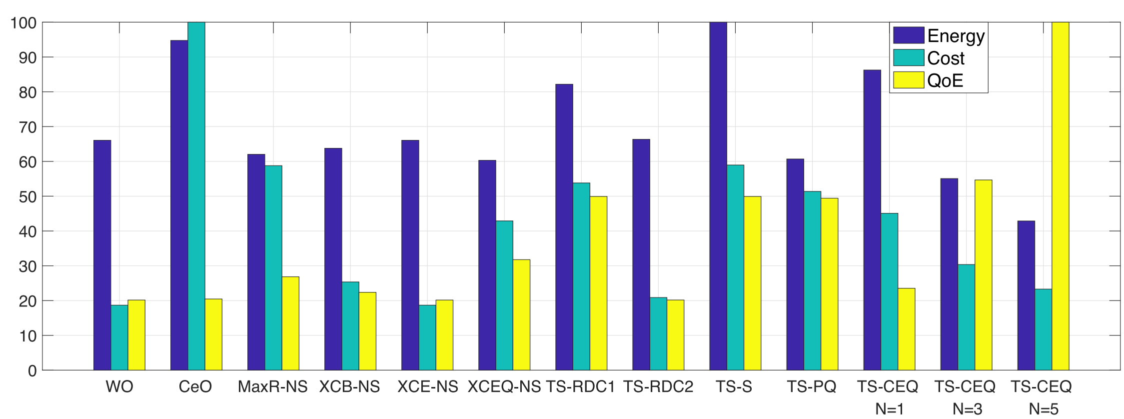


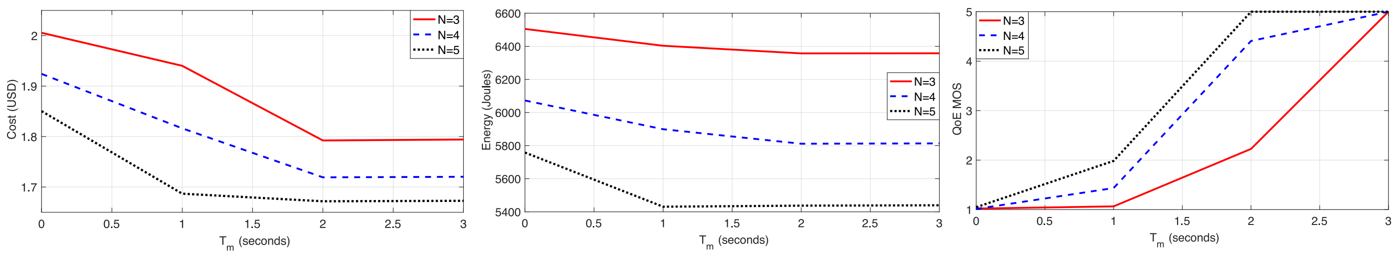
Fig. 7. Comparison of total energy expenditure, monetary cost and QoE MOS performance (%) with respect to the strategies providing highest energy consumption (TS-S), highest monetary cost (CeO) and highest QoE MOS (TS-CEQ with $N = 5$)

5.3. Simulations results and analysis

To compare the performance of the various strategies mentioned in Section 5.1, we evaluated the total energy consumption, monetary cost and QoE for one hour of video streaming, with an arrival rate of 117 kBytes every second, considering two case scenarios. In the first case scenario, we considered only two wireless interfaces, WiFi and cellular networks. We analyzed the performance of the proposed TS-CEQ under different system parameters such as the number of time slots N, minimum required buffered video in seconds T_m, and cost weight factor α. In the second case, we tested the performance of the proposed approach considering device-to-device communication and using more than two interfaces simultaneously for data download.

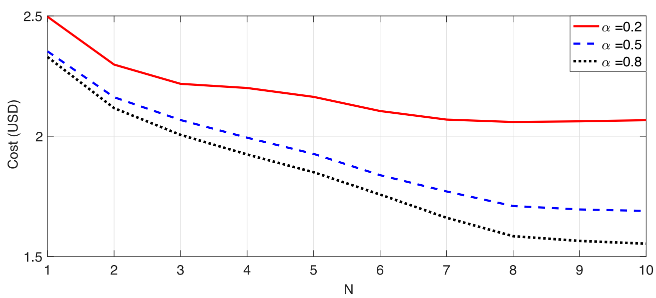
5.3.1. Results considering two interfaces – WiFi and cellular networks

In the first scenario, the WiFi and cellular transmission rates are modeled following an exponential distribution with average rate of 150 kBps. To compare the performance of the various approaches presented in Section 5.1, we quantified the tradeoff between the monetary cost, the energy consumption and the QoE mean opinion score. We assume α to be 0.8, giving higher weight to the monetary cost and T_m to be 0 with no requirement on minimum video buffered data. The effect of α and T_m on the performance of the proposed TS-CEQ approach is presented in Section 5.3.2.

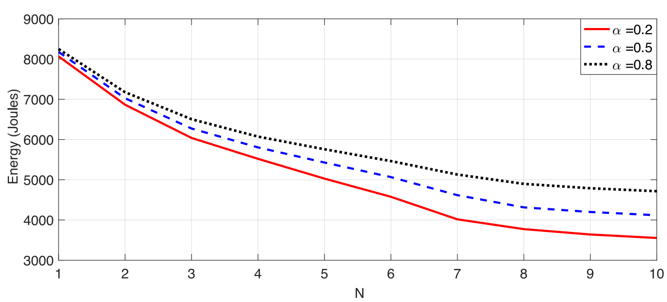


(a) Cost (USD) (b) Total energy consumption (Joules) (c) QoE mean opinion score

Fig. 8. TS-CEQ performance in terms of monetary cost, energy consumption and QoE while varying N and T_m with α=0.8.



(a) Cost (USD)



(b) Total energy consumption (Joules)

Fig. 9. TS-CEQ performance in terms of monetary cost and energy consumption while varying N and α with T_m=0.

Fig. 5 presents the tradeoff between total energy consumption in Joules versus the monetary cost in USD for every approach. Fig. 6 presents the total energy consumption versus the QoE mean opinion score. In addition, since the aim is to minimize the energy expenditure and monetary cost while increasing user quality of experience, Fig. 7 presents the percentages of performance evaluation for every approach in terms of energy, cost and QoE compared to the strategies providing highest energy consumption (TS-S), highest monetary cost (CeO) and highest QoE MOS (TS-CEQ with $N = 5$), respectively.

The results show low performance for network selection strategies including WiFi only, cellular only, maximum rate network selection, XCB-NS, XCE-NS and XCEQ-NS based on [9] strategies. Using WO, with cost 0.003×10^{-3} USD per packet and low receiving power of 1.307 W, provides low energy consumption and monetary cost at the expense of quality of experience, which was the lowest with MOS 1.008. Using CeO provides low QoE with high energy consumption and monetary cost. XCB-NS provides lower energy and monetary cost compared to MAXR-NS. This is due to the fact that XCB-NS aims at selecting the network providing a tradeoff between low cost and high bandwidth or rate, whereas MAXR-NS selects the network providing only higher transmission rate. We customized the work in [9] to consider XCE-NS aiming at selecting the interface providing minimum energy and monetary cost, which leads to the solution of selecting WiFi only, without taking QoE into consideration. We also customized their work and considered a network selection approach providing monetary cost-energy-QoE tradeoff. XCEQ-NS provides higher QoE compared to other network selection approaches XCE-NS, XCB-NS and MAXR-NS with lower energy consumption while consuming more monetary cost.

Using one interface for data download is not sufficient to provide the user with a download rate greater than the arrival rate, which leads to a low QoE level. The approaches where traffic splitting is considered provided better performance in terms of user satisfaction, energy consumption and cost. The user takes advantage of downloading data over different wireless interfaces simultaneously to maintain a high download rate, which decreases the stalling events, their frequency and length, hereby, enhances user satisfaction.

The TS-S approach forces the data to be downloaded over both interfaces simultaneously, which leads to maximizing the instantaneous data rate at every time slot without considering energy consumption minimization, nor QoE level maximization. TS-PQ presented in [14] aims at providing a balance between QoE, energy consumption and delay while considering one time slot ahead for future prediction. TS-PQ provides a QoE MOS of 2.47, which is very close to the TS-S approach when using both links simultaneously, while consuming approximately 40% less energy, and 13% less monetary cost. This is due to the fact that TS-PQ may choose to defer transmission or select one of the interfaces, if the playing buffer is full or the data transmission rate is low, leading to a lower monetary cost, energy consumption and delay while keeping a high QoE. The traffic splitting approach TS-RDC proposed in [15] aims at providing a balance between link capacity, delay and monetary cost. We consider TS-RDC1 where the weights in the utility function of the link capacity, delay and monetary cost are 0.2, 0.5, and 0.3, respectively, as presented in [15]. In TS-RDC2, we customize their work and gave higher weight for minimizing the monetary cost to match our objective function. TS-RDC1 consumes 26% more energy than TS-PQ, with very close QoE MOS and monetary cost. This is due to the fact that both traffic splitting approaches aim at enhancing user satisfaction; TS-RDC1 gives very high weight to delay and link capacity, and TS-PQ maximizes instantaneous QoE, while also considering energy consumption. Giving higher weight to monetary cost, TS-RDC2 provides 59% lower monetary cost compared to TS-PQ, at a tradeoff cost in quality of experience and energy consumption. In order to minimize the cost, TS-RDC2 uses the WiFi interface more frequently which affects the QoE and leads to a solution very close to using WiFi alone.

When the proposed approach TS-CEQ considers only one time slot ahead (N=1), the cost was lower than TS-PQ and TS-S by 12.2% and 23.5%, respectively with lower QoE of 1.17. The energy consumed was 40% lower than TS-S and 30% higher than TS-PQ. This is due to the fact that TS-PQ aims at minimizing energy and maximizing QoE, whereas TS-CEQ considers mainly minimizing the monetary cost and energy consumption while maintaining a minimum QoE bound. Considering the network performance for multiple time slots ahead enhanced the proposed approach performance. When $N = 3$ time slots ahead were considered for rate estimation, TS-CEQ consumed 40.8% less cost, 9.2% less energy while providing 10% better quality of experience when com-

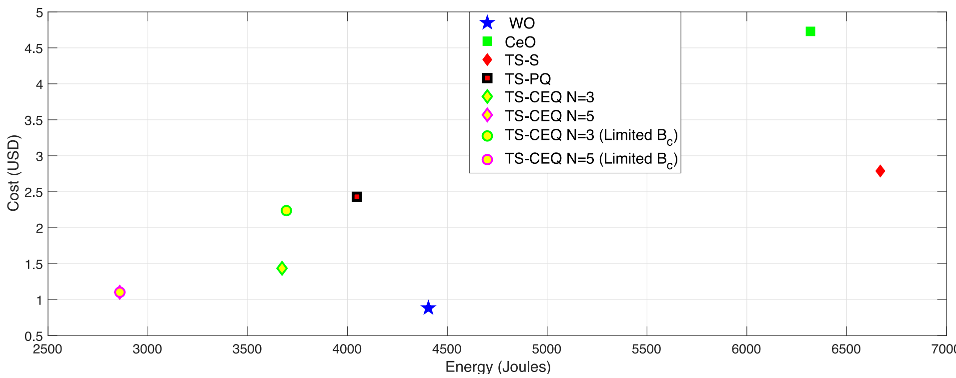


Fig. 10. Total energy expenditure (Joules) and monetary cost (USD) for every approach.

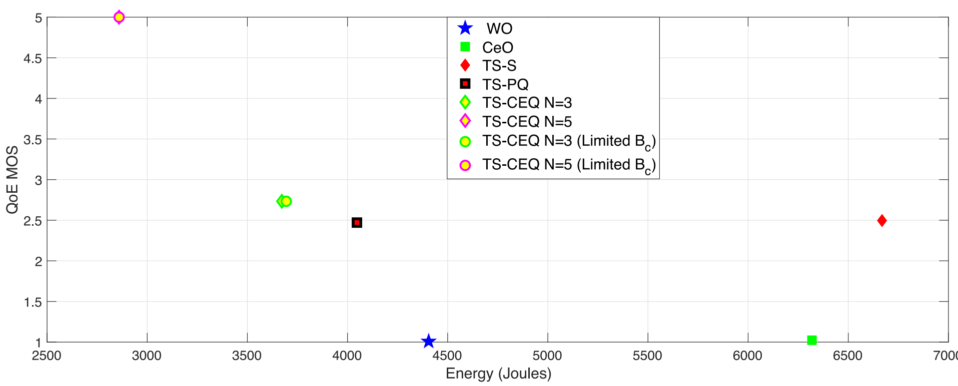


Fig. 11. Total energy expenditure (Jo ules) and QoE MOS for every approach.

pared to TS-PQ. TS-CEQ with $N = 5$ outperformed all other algorithms, while providing an excellent QoE of 5, with more than 57% reduction in energy consumption compared to TS-S and 76.6% (3.62 USD) reduction in monetary cost compared to CeO. Compared to TS-PQ, TS-CEQ with $N = 5$ provides 54.63% (1.32 USD) reduction in monetary cost, 29.34% reduction in energy consumption, and 50.5% increase in user satisfaction. Increasing the considered number of time slots N ahead for future prediction increases the system performance in terms of cost, energy and QoE since the optimal decision will have a clearer futuristic vision of the channel estimation for a higher number of time slots. Accordingly, the TS-CEQ proposed approach may defer transmission at a time slot t to a future time slot j with $j = [t, t+1, \ldots, t+N-1]$, where the transmission rate is higher, leading to a lower delay and energy consumption while maintaining QoE level and guaranteeing the minimum number of data packets buffered.

5.3.2. *Study on the TS-CEQ system parameters* – N, *α and* T_m

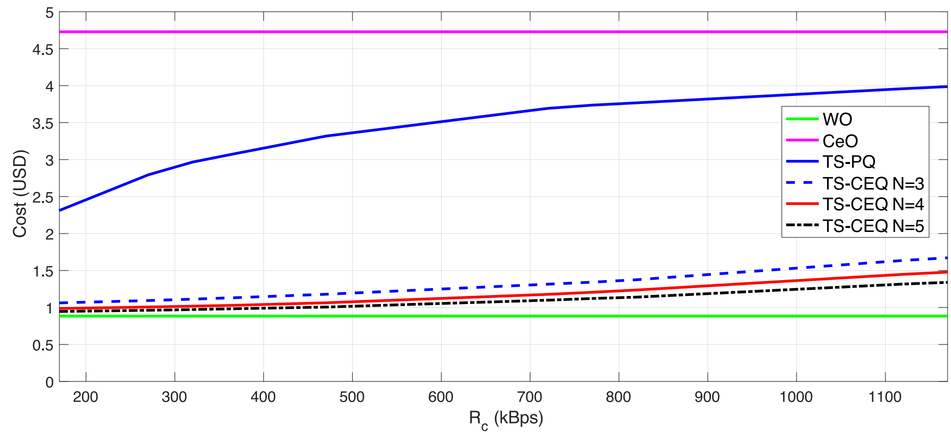
In this section, we evaluate the performance of the proposed TS-CEQ approach while varying system parameters such as the number N of time slots considered for future prediction, the weighting factor α and the required minimum time for video data buffering T_{m}. We assume the network is composed of WiFi and cellular networks, and the transmission rates are modeled following an exponential distribution with average rate of 110 kBps considering video streaming for one hour. Fig. 8 shows the monetary cost in USD, the energy consumption in Joules and the QoE mean opinion score provided by the TS-CEQ approach for N=3, 4 and 5, while varying T_{m} from 0 to 3 s. In this case, α is fixed to 0.8 to give higher weight to the monetary cost with respect to energy consumption. Results prove that considering higher number of time slots N ahead for future prediction increases the system performance in terms of cost, energy and QoE. Considering larger number of N allows to take advantage of the estimation of the channel performance for more time slots ahead. This will allow the proposed approach to defer transmission at a specific time slot to benefit from a better link performance at later time slots, leading to a lower delay and energy consumption while maintaining QoE level. For instance, considering N=5 instead of 3 was able to reduce the cost and the energy by more than 6.7% and 11.5%, respectively, while providing higher quality of experience reaching MOS of 5 (double the QoE when N=3) when T_{m}=2 s. On the other hand, increasing T_{m} provides slightly lower monetary cost and energy consumption while providing higher level of QoE. Increasing the minimum buffered time of video streaming T_{m} requirement prevented the buffer queue to go below D_{m} and consequently prevented stalling events. As shown in Fig. 8, increasing T_{m} has major impact on QoE which increases from 1 (poor quality) to 5 (excellent quality) when T_{m} is greater than 2 s.

Fig. 9 shows the monetary cost in USD and the energy consumption in Joules provided by the TS-CEQ approach for α=0.2, 0.5 and 0.8, while varying N from 1 to 10. In this case, T_{m} is fixed to 0. α is a weighting factor indicating the tradeoff between cost and energy consumption; i.e. increasing α gives higher impact to reducing the monetary cost instead of energy consumption. The results in Fig. 9 show that increasing α from 0.2 to 0.8 is able to reduce the download cost while increasing the energy consumption which is align with our proposed objective function.

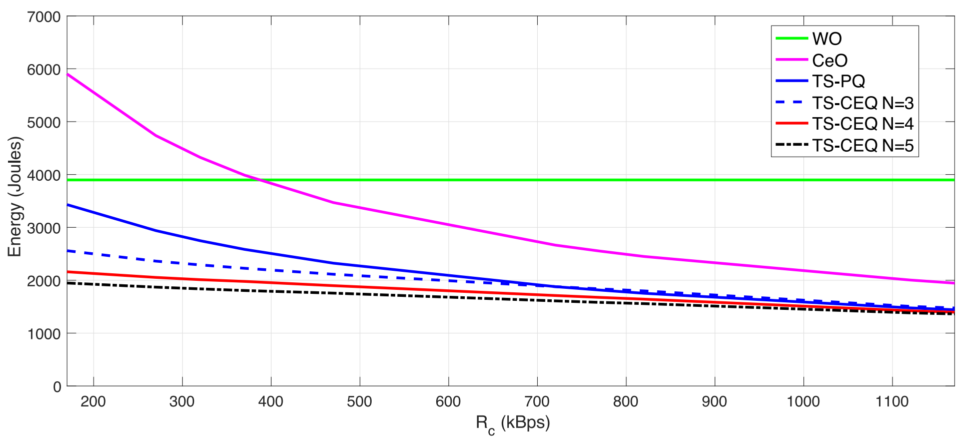
5.3.3. *Study on the TS-CEQ performance with limited cellular budget*

Figs. 10 and 11 compare the performance of the TS-CEQ approach with/without limited cellular budget. In this scenario, we consider a limited cellular budget of 50MB, with $\theta_2 = 0.016 \times 10^{-3}$ USD/packet (if below the budget) and $\Gamma_2 = 0.116 \times 10^{-3}$ USD/packet (if above the budget). We assume the WiFi and cellular transmission rates are modeled following an exponential distribution with average rate of 150 kBps for one hour video streaming.

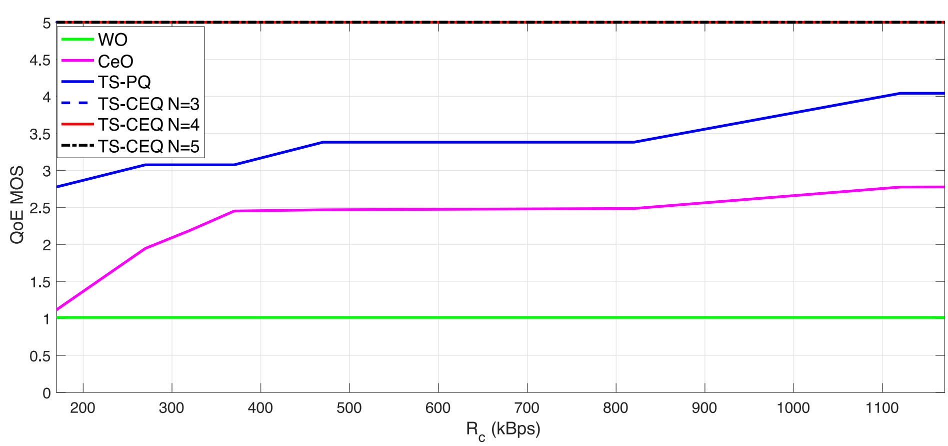
The TS-CEQ approach is able to provide the same QoE level while consuming more energy and monetary cost for $N = 3$. Due to the limitation in the cellular data budget, the mobile is encouraged to use the WiFi



(a) Cost (USD)



(b) Total energy consumption (Joules)



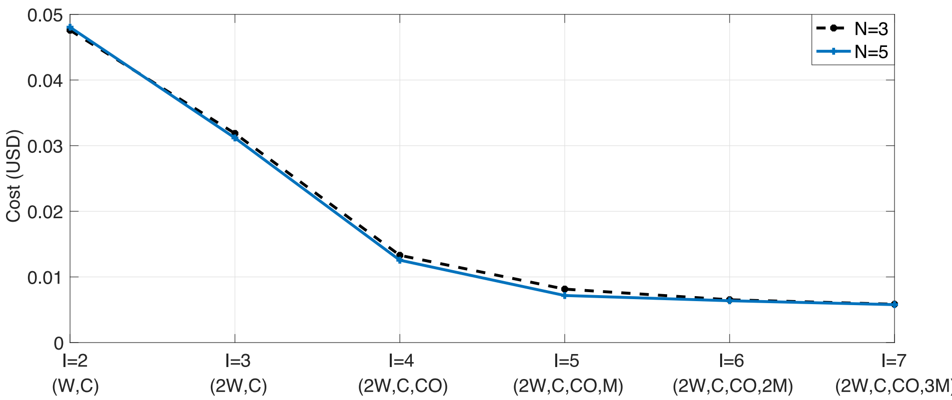
(c) QoE mean opinion score

Fig. 12. TS-CEQ performance in terms of monetary cost, energy consumption and QoE while varying average cellular rate R_c.

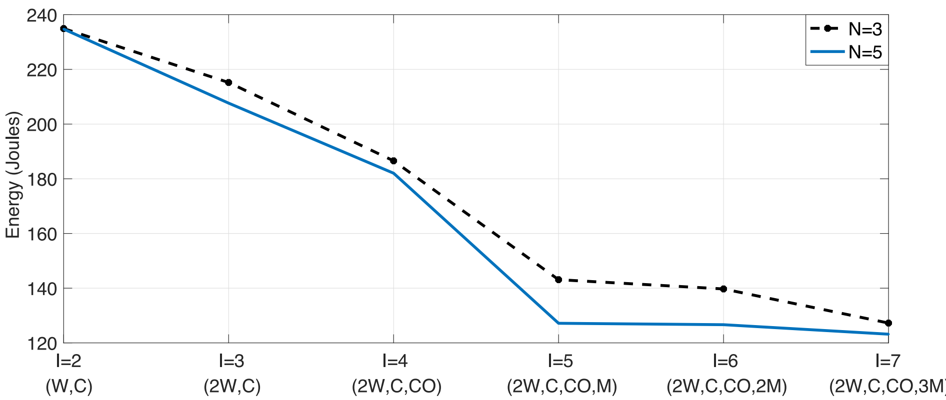
network more often even with low WiFi rates to satisfy the minimum QoE requirement. In addition, the cost of data packet is higher when the user is above cellular data budget. This leads to higher energy and cost consumption in order to maintain the minimum user satisfaction level. For $N = 5$, the TS-CEQ approach was able to overcome the budget limitation by providing similar performance and keeping the monetary cost low. The WiFi and cellular interfaces were used intelligently in order to conserve the budget and use the WiFi interface more often while satisfying the minimum QoE level requirement.

5.3.4. *Study on the TS-CEQ performance with different cellular average rates*

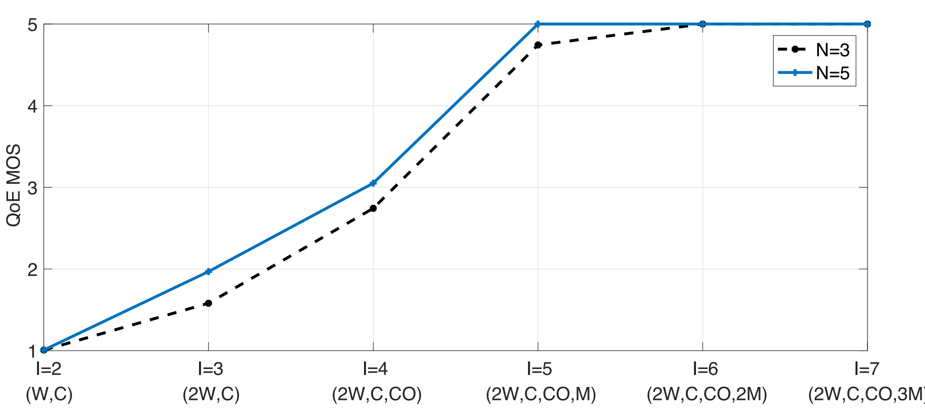
In the previous scenarios, we used symmetrical rates for both WiFi and cellular links. To show the performance of more realistic scenarios, we compared the performance of asymmetrical WiFi and cellular trans-



(a) Cost (USD)



(b) Total energy consumption (Joules)



(c) QoE mean opinion score

Fig. 13. TS-CEQ performance in terms of monetary cost, energy consumption and QoE while varying the number of wireless interfaces.

mission rates considering video streaming for one hour duration. In our simulations, the WiFi transmission rate follows an exponential distribution with average of 200 kBps. The cellular rate follows the exponential distribution with average transmission rate R_c varying from 200 kBps to 1.2 MBps. We assume α and T_m are set to 0.8 and 0, respectively.

Fig. 12 (a)–(c) evaluate the performance of the different approaches in terms of monetary cost, energy consumption and QoE MOS, respectively. The results in Fig. 12(a) show that by increasing R_c, the monetary cost slightly increases since the approaches tend to use the cellular interface more often to take advantage of the higher rate compared to WiFi and achieve performance gains in terms of transmission time, energy consumption and QoE. In Fig. 12(b), we evaluate the transmission energy consumption consumed by the mobile device to download the data over every interface. When R_c increases, the transmission time decreases. Accordingly, the energy consumption decreases with the increase of the average cellular rate. As presented in Fig. 12(c), increasing the average transmission cellular rate provides higher QoE. The interfaces data capacity satisfied the video arrival rate and needed number of packets which reduced the stalling events, their frequency and length, and hereby, increasing the user satisfaction. The proposed approach TS-CEQ was able to provide a maximum QoE level of 5. Using WiFi only provided a MOS of 1 with an average rate of 200 kBps, while using cellular alone, the QoE MOS values increased to reach 2.77 at an average rate of 1.2 MBps.

5.3.5. Results considering multiple wireless interfaces – WiFi and cellular networks with D2D cooperation

In order to assess the performance of the use of multiple wireless interfaces simultaneously, we considered the co-existence of the cellular network with multiple WiFi-hotspots, content owners mobile devices, and multiple cooperative devices. In this scenario, we assume the transmission rate for every interface follows the exponential distribution with a low rate of 40 kBps. We evaluated the monetary cost, energy consumption, and QoE for 1 min of video streaming while varying the number of interfaces. Fig. 13 shows the performance of 6 different data transmission scenarios with different number of interfaces as follows: (1) $I = 2$ (W,C) where two interfaces WiFi and cellular are available to the user, (2) $I = 3$ (2W,C) with three interfaces two WiFi hotspots and cellular network, (3) $I = 4$ (2W,C,CO) with two WiFi hotspots, cellular network and a content owner, (4) $I = 5$ (2W,C,CO,M), (5) $I = 6$ (2W,C, CO,2M) and (6) $I = 7$ (2W,C,CO,3M) with five, six and seven available interfaces including two WiFi hotspots, cellular network, a content owner, and one, two and three cooperative mobile devices, respectively, connected to a WiFi hotspot charged with 0.003×10^{-3} USD/packet each. The results show that increasing the number of interfaces was able to enhance system performance in terms of increasing QoE while reducing monetary cost and energy consumption.

The existence of higher number of interfaces allows the mobile terminal to achieve higher rate while downloading data over multiple wireless interfaces simultaneously. This has high impact on reducing the transmission time and delay while avoiding data buffer underflow and stalling events. The quality of experience increased from 1 (poor) when WiFi and cellular interfaces are used to 5 (excellent) when 5 interfaces are used. In our work, we consider the energy consumed for transmission by the mobile devices for data download. The reduction in energy consumption is due to the reduction of the interfaces utilization and transmission time, in addition to the use of D2D communication. To download the same video, using seven interfaces simultaneously including D2D cooperation provided 47.52% reduction in energy consumption, 87.91% reduction in monetary cost while achieving an excellent QoE level of 5, compared to the use of WiFi and cellular interfaces.

5.4. Simulations results outcome

The proposed TS-CEQ approach provides dynamic optimal solutions aiming at minimizing the energy consumption and monetary cost, while ensuring high QoE by avoiding buffer underflow and stalling events. The latter is achieved by simultaneously utilizing the multiple wireless interfaces available at the mobile terminal. Using bit rate estimation for the subsequent N time slots allowed TS-CEQ to outperform other network selection and traffic splitting approaches. TS-CEQ with $N = 5$ even with limited cellular budget was able to provide an excellent QoE of 5, with more than 57% and 29.34% reduction in energy consumption compared to TS-S and TS-PQ, respectively, and more than 76.6% (3.62 USD) and 54.63% (1.32 USD) reduction in monetary cost compared to CeO and TS-PQ, respectively. The user may achieve excellent quality level even with very low transmission rates using five interfaces, in addition to lower transmission energy consumption and monetary cost.

6. Conclusions

This paper addressed user-centric real-time traffic splitting in D2D cooperative heterogeneous networks where a mobile can take advantage of multiple wireless interfaces simultaneously to reduce energy consumption, delay, monetary cost while maintaining a minimum quality of experience level. We formulated the problem as a multi-objective optimization problem with cost-energy- QoE tradeoffs. We considered different cost modeled and allowed futuristic time slot rate estimation. Results demonstrate significant performance gains compared to conventional networks with an excellent QoE of 5 with more than 57% reduction in energy consumption and 76% reduction in monetary cost, with very low solving time.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Cisco, Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2017–2022, White Paper, Cisco, 2019.

[2] 3GPP TS 22.278 version 8.5.0 Release 8, Service requirements for the Evolved Packet System (EPS) (2008).

[3] 3GPP TR 36.839 version 11.0.0 Release 11, Evolved Universal Terrestrial Radio Access (E-UTRA); Mobility enhancements in heterogeneous networks (2012).

[4] X. Li, Q. Huang, D. Wu, A repeated stochastic game approach for strategic network selection in heterogeneous networks, in: Proceedings of the IEEE INFOCOM - IEEE Conference on Computer Communications Workshops, 2018.

[5] T. ZhouN. Jiang, Z. Liu, Zunxiong and C. Li, T. Zhou, N. Jiang, Z. Liu, Zunxiong, C. Li, Joint cell activation and selection for green communications in ultra-dense heterogeneous networks, IEEE Access 6 (2018) 1894–1904.

[6] J. Montoya, A. Sethi, N.G. Gómez, A load-based and fair radio access network selection strategy with traffic offloading in heterogeneous networks, in: Proceedings of the 7th International Conference on Computers Communications and Control, 2018.

[7] X. Wang, J. Li, L. Wang, C. Yang, Z. Han, Intelligent user-centric network selection: a model-driven reinforcement learning framework, IEEE Access 7 (2019) 21645–21661.

[8] G. Araniti, P. Scopelliti, G. Muntean, A. Iera, A hybrid unicast-multicast network selection for video deliveries in dense heterogeneous network environments, IEEE Trans. Broadcast. 65 (1) (2019) 83–93.

[9] S. Khan, M.I. Ahmad, F. Hussain, Exponential utility function based criteria for network selection in heterogeneous wireless networks, Electron. Lett. 54 (8) (2018) 529–531.

[10] P. Zhao, W. Yu, X. Yang, D. Meng, L. Wang, Buffer data-driven adaptation of mobile video streaming over heterogeneous wireless networks, IEEE Internet Things J. 5 (5) (2018) 3430–3441.

[11] P. Go, OC. Kwon, H. Song, An energy-efficient HTTP adaptive video streaming with networking cost constraint over heterogeneous wireless networks, IEEE Trans. Multimed. 17 (9) (2015) 1646–1657.

[12] K. Nguyen, M.G. Kibria, J. Hui, K. Ishizu, F. Kojima, Minimum latency and optimal traffic partition in 5G small cell networks, in: Proceedings of the 87th IEEE Vehicular Technology Conference, 2018.

[13] S. Sharafeddine, K. Jahed, M. Fawaz, Optimized device centric aggregation mechanisms for mobile devices with multiple wireless interfaces, Elsevier Comput. Netw. 129 (2017) 1–16.

[14] N. Abbas, H. Hajj, Z. Dawy, K. Jahed, S. Sharafeddine, An optimized approach to video traffic splitting in heterogeneous wireless networks with energy and QoE considerations, Elsevier J. Netw. Comput.Appl. 83 (2017) 72–88.

[15] A.I. Abdulshakoor, M.M. Elmesalawy, N.A. Elmosilhy, A.M.A El-Haleem, Joint network and mode selection in 5G multi RAT heterogeneous networks, in: Proceedings of the 42nd International Conference on Telecommunications and Signal Processing, 2019.

[16] A.I. Abdulshakoor, M.M. Elmesalawy, G.A. Khalaf, Proportional traffic splitting for efficient LTE-WLAN aggregation in multi-RAT heterogeneous networks, in: Proceedings of the 36th IEEE National Radio Science Conference, 2019.

[17] J.J. Nielsen, R. Liu, P. Popovski, Ultra-reliable low latency communication using interface diversity, IEEE Trans. Commun. 66 (3) (2018) 1322–1334.

[18] N. Abbas, S. Sharafeddine, H. Hajj, Z. Dawy, Cost and energy aware dynamic splitting of video traffic in heterogeneous networks, in: Proceedings of the 24th IEEE symposium on Computers and Communications, 2019.

[19] G. Feng, F. Xia, Y. Zhang, D. Su, H. Lv, H. Wang, Optimal cooperative wireless communication for mobile user data offloading, IEEE Access 6 (2018) 16224–16234.

[20] M. Shih, T. Wu, H. Wei, Unlicensed LTE pricing for tiered content delivery and heterogeneous user access, IEEE Trans. Mob. Comput. 18 (1) (2019) 235–249.

[21] S. Noreen, N. Saxena, A. Roy, Discount interference pricing mechanism for data offloading in D2D communications, IEEE Commun. Lett. 22 (8) (2018) 1688–1691.

[22] W. Bai, X. Gan, X. Wang, X. Di, J. Tian, Optimal data traffic pricing in socially-aware network: a game theoretic approach, in: Proceedings of the IEEE Global Communications Conference, 2016.

[23] H. Chen, L. Xiao, D. Yang, T. Zhang, L. Cuthbert, User cooperation in wireless powered communication networks with a pricing mechanism, IEEE Access 5 (2017) 16895–16903.

[24] Y. Chen, S. He, F. Hou, A pricing strategy for D2D communication from a prospect theory perspective, in: Proceedings of the 89th IEEE Vehicular Technology Conference, 2019.

[25] V. Passas, V. Miliotis, N. Makris, T. Korakis, Dynamic RAT selection and pricing for efficient traffic allocation in 5G HetNets, in: Proceedings of the IEEE International Conference on Communications, 2019.

[26] M. Jin, S. Gao, H. Luo, J. Li, Y. Zhang, S.K. Da, An approach to pre-schedule traffic in time-dependent pricing systems, IEEE Trans. Netw. Serv.Manag. 16 (1) (2019) 334–347.

[27] S. Sen, C. Joe-Wong, S. Ha, M. Chiang, Smart Data Pricing (SDP): economic solutions to network congestion, in: H. Haddadi, C. Bonaventure (Eds.), Recent Advances in Networking, 2013, pp. 221–274.

[28] S. Sen, C. Joe-Wong, S. Ha, M. Chiang, Incentivizing time-shifting of data: a survey of time-dependent pricing for Internet Access, IEEE Commun. Mag. 50 (11) (2012) 91–99.

[29] S. Sharafeddine, K. Jahed, N. Abbas, E. Yaacoub, Z. Dawy, Exploiting multiple wireless interfaces in smartphones for traffic offloading, in: Proceedings of the First International Black Sea Conference on Communications and Networking, 2013.

[30] N. Abbas, S. Taleb, H. Hajj, E. Yaacoub, Z. Dawy, A learning-based approach for network selection in WLAN/3G heterogeneous network, in: Proceedings of the Third International Conference on Communications and Information Technology, June 2013.

[31] K. Jahed, M. Younes, S. Sharafeddine, E. Yaacoub, Energy measurements for mobile cooperative video streaming, in: Proceedings of the IFIP Wireless Days, 2012.

[32] [ITU-T P.1201] recommendation ITU-T P.1201 (2013). Parametric Non-Intrusive Assessment of Audiovisual Media Streaming Quality - Amendment 2: New Appendix III — Use of ITU-T P.1201 for Non-Adaptive, Progressive Download Type Media Streaming, December 2013.

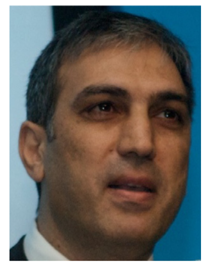
[33] ITU-T P.10/G.100, Recommendation ITU-T P.10/G.100 (2016), Amendment 1: New Appendix I — Definition of Quality of Experience (QoE), 2006.

[34] ITU-T P.1202, Recommendation ITU-T P.1202 (2012), Parametric Non-Intrusive Assessment of Audiovisual Media Streaming Quality, 2012.

[35] European Telecommunications Standards Institute, ETSI TR 102 714 V1.1.1. Speech and Multimedia Transmission Quality (STQ); Multimedia Quality Measurement; End-to-End Quality Measurement Framework 2011.

Nadine Abbas is a lecturer in the Electrical and Computer Engineering department and research fellow in the Computer Science department at the Lebanese American University American University (LAU). Nadine earned a Bachelor of Engineering degree in Computer and Communication Engineering from Notre Dame University in 2008 with highest distinction. She received her M.E. and Ph.D. degrees in Electrical and Computer Engineering from American University of Beirut (AUB) in 2011 and 2017, respectively. Her research interests include wireless communications, mainly heterogeneous networks that support intelligent multi-radio networking.

Sanaa Sharafeddine is an Associate Professor of Computer Science at the Lebanese American University. She received the B.E. and the M.E. degrees in Computer and Communications Engineering from the American University of Beirut in 1999 and 2001, respectively. She received her doctoral degree in Communications Engineering from Munich University of Technology (TUM) in 2005 in collaboration with Siemens AG research labs in Munich. She received L'Oreal-UNESCO's International Rising Talent Award in 2015 and Pan-Arab Regional Fellowship Award in 2013. She currently serves on the editorial boards of IEEE Journal on Selected Areas in Communications' Series on Network Softwarization and Enablers, IEEE Access, and Elsevier Ad Hoc Networks. Her research interests include UAV-assisted solutions in 5G, wireless networking, and multimedia services.



Hazem Hajj is an Associate Professor with the American University of Beirut (AUB) and a senior member of IEEE. Hazem joined AUB in 2008, and was a visiting Associate Professor at the University of Texas-Austin for one year in 2015-16. Before joining AUB, Hazem was a principal engineer at Intel Corporation. He received his PhD from the University of Wisconsin-Madison in 1996, and his Bachelor from AUB with distinction. Hazem's research interests include Machine Learning and Energy-Aware Computing, with special interests in Context Aware Sensing and Natural Language Understanding. He has over 100 research publications in reputable journals and international conferences.

Zaher Dawy received the B.E. Degree in Computer and Communications Engineering from American University of Beirut (AUB) in 1998 and the Doctoral Degree in Communications Engineering from Munich University of Technology (TUM) in 2004. He joined AUB in 2004 and is currently a Professor in the Electrical and Computer Engineering Department. His research interests are in the general areas of wireless communications and biomedical engineering. He currently serves on the Editorial Board of IEEE Transactions on Communications, IEEE Transactions on Wireless Communications, IEEE Communications Surveys and Tutorials, and Computer Networks (Elsevier).